IJoL International Journal of Librarianship

ISSN: 2474-3542 Journal homepage: http://journal.calaijol.org

Linked Data Technologies and What Libraries Have Accomplished So Far

Yongming Wang and Sharon Q. Yang

Abstract:

For the past ten years libraries have been working diligently towards Linked Data and the Semantic Web. Due to the complexity and vast scope of Linked Data, many people have a hard time to understand its technical details and its potential for the library community. This paper aims to help librarians better understand some important concepts by explaining the basic Linked Data technologies that consist of Resource Description Framework (RDF), the ontology, and the query language. It also includes an overview of the achievements by libraries around the world in their efforts to turn library data into Linked Data including those by Library of Congress, OCLC, and some other national libraries. Some of the challenges and setbacks that libraries have encountered are analyzed and discussed. In spite of the difficulties, there is no way to turn back. Libraries will have to succeed.

To cite this article:
Wang, Y., & Yang, S. Q. (2018). Linked Data Technologies and What Libraries Have Accomplished So Far. *International Journal of Librarianship, 3*(1), 3-20. doi: https://doi.org/10.23974/ijol.2018.vol3.1.62

INTERNATIONAL JOURNAL OF LIBRARIANSHIP, 3(1), 3-20
ISSN:2474-3542

Linked Data Technologies and What Libraries Have Accomplished So Far

Yongming Wang, The College of New Jersey

Sharon Q. Yang, Rider University

ABSTRACT

For the past ten years libraries have been working diligently towards Linked Data and the Semantic Web. Due to the complexity and vast scope of Linked Data, many people have a hard time to understand its technical details and its potential for the library community. This paper aims to help librarians better understand some important concepts by explaining the basic Linked Data technologies that consist of Resource Description Framework (RDF), the ontology, and the query language. It also includes an overview of the achievements by libraries around the world in their efforts to turn library data into Linked Data including those by Library of Congress, OCLC, and some other national libraries. Some of the challenges and setbacks that libraries have encountered are analyzed and discussed. In spite of the difficulties, there is no way to turn back. Libraries will have to succeed.

Keywords: Linked Data, Semantic Web, Resource Description Framework, BIBFRAME, Library of Congress, OCLC

INTRODUCTION

What is Linked Data? According to David Wood, the co-chair of the W3C's (World Wide Web Consortium) RDF Working Group which lays the foundation for Linked Data and the Semantic Web, "Linked Data is a set of techniques to represent and connect structured data on the web… Linked Data makes the World Wide Web into a global database that we call the Web of Data" (Wood, Zaidman, Ruth, & Hausenblas, 2014). Linked Data technologies, with its broader concept, Semantic Web, has gained rapid momentum and popularity on the World Wide Web. The Linked Data technologies hold the potential to evolve the current Web of document into the Web of Data. Imagine that in the future Internet world, not only web documents but all data are

connected. More importantly, these connected data are not only accessible to human but to machine also. In other words, all devices that are connected on the Internet can access and process those linked data and thereby make smart decisions automatically. This will greatly enhance the way we access information and make informed decisions. The ideas are not new. As early as late 90's, Tim Berners-Lee (2000), the inventor of World Wide Web, had a vision for Semantic Web:

> "I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize."

In 2004, W3C published the first recommendation of the data model for Linked Data, the RDF 1.0. In 2005, W3C formed the Semantic Web Interest Group. And in 2006, Tim Berners-Lee published the Linked Data principles and design rules, which paves the way for large scale adoption and development of Linked Data technologies (Berners-Lee, 2006).

The last ten plus years has witnessed rapid adoption and usage of Linked Data by companies small and large. Companies such as Google and Facebook use Linked Data to enhance their searching capability and connections (Wood et al., 2014). Retail company BestBuy uses Linked Data to improve its business bottom lines (Wood et al., 2014).

LITERATURE REVIEW

As early as 2011, the Library Linked Data Incubator Group (2011) published its final report as a W3C Incubator Group Report. This group consisted of the international experts in the library and information fields who are specialized in Semantic Web and metadata. In this report, it surveyed the current situation of Linked Data, summarized the use cases, and made some important recommendations for implementing the Linked Data in the library community.

Another international effort that closely relates to Semantic Web and Linked Data is the annual international conference on Dublin Core and Metadata Applications by the Dublin Core Metadata Initiative (DCMI). This annual conference started in 1995 as workshops only and in 2001 expanded to full conferences with additions of tutorials, presentations, and peer-reviewed papers. From the early on, DCMI tackles the issues related to Semantic Web, especially the ontology and vocabularies. The theme of 2005 conference is "Vocabularies in Practice." One paper in this year's proceeding introduced the concept of SKOS (Simple Knowledge Organization System) and recommended a way to use SKOS Core and DCMI Metadata Terms in combination (Miles, 2005). One project report in the 2009 conference proceeding has the title "Research on Linked Data and Co-reference Resolution," which described the transformation of a dataset of academic authors and their publications into Linked Data (Glaser, 2009). This is one of the earliest publications on Linked Data application in library community. And since 2012,

there has been increasing focus on the topic of Linked Data in this annual conference series.

In more practical area, Karen Coyle (2012) published "Linked Data Tools: Connecting on the Web" in Library Technology Report. In this report, she introduced the basic technologies of Linked Data in a tutorial format. A year later, Erik T. Mitchell (2013) published "Library Linked Data: Research and Adoption" in Library Technology Report, in which he talked about the development and research of Linked Data in library community. In 2016, Mitchell (2016) published another report dealing with the library adoption and practice of Linked Data entitled "Library Linked Data: Early Activity and Development." The three reports by Colye and Mitchell have played an important role in helping librarians learn about Linked Data.

Since 2015, more articles on the case studies and use examples in Linked Data have been published. Karim Tharani's article (2015) explores the possibility of using BIBFRAME to harvesting and sharing bibliographic data as linked data by ways of a case study. The article of Jin, Hahn, and Croll (2016) also talks about their project of transforming and enriching nearly 300,000 e-books MARC records to BIBRAME records and in the meantime increasing the discoverability of accessibility of those resources. Kimmy Szeto's article explores how linked open data can transform and enhance the discovery and search of music resources. (Szeto, 2017) Recently, another project by OCLC's PCC (Program for Cooperative Cataloging) was carried out to transform the legacy library metadata, that is, the MARC records, to Linked Data. The major task of this project is to create a Linked Data authority control database by aggregating the traditional MARC records of people, organization, and places from many sources and converting them into Linked Data. As stated in PCC 2015-2017 strategic directions (Godby & Smith-Yoshimura, 2017):

> "Existing methods of library authority control are based on constructing unique authorized access points as text strings (literals). This string-based approach works somewhat well in the closed environment of a traditional library catalog, but not in an open environment where data are shared and linked, and so require unique identifiers. The web presents both a challenge and an opportunity for libraries, which are now in a position to take advantage of data created outside of the library world, and also to contribute library authority data for use by other communities" (p.20).

Another issue in transforming library legacy metadata into Linked Data is that there are several efforts from different library organizations, resulting in different conceptual models. Zepounidou et al. (2017) compare four conceptual models, namely Functional Requirements for Bibliographic Records (FRBR), FRBR Object-Oriented (FRBRoo), Bibliographic Framework (BibFrame), and Europeana Data Model (EDM), and try to find the common ground and convergences among them. Therefore, the goal of interoperability can be realized.

There are many publications in Linked Data. But sometimes the librarians still feel it's a challenge to understand the concept of Linked Data. According to Banerjee (2017): "Even though librarians have read about and attended sessions discussing topics such as linked data and FRBR (Functional Requirements for Bibliographic Records) for more than a decade, they still find these things confusing." (p.21)

LINKED DATA TECHNOLOGY

Resource Description Framework (RDF) Data Model

Simply put, a data model is an abstract of real data and their relationships. The most familiar data model we encounter is the tabular data model such as csv file, which lists data in table structured format.

The data model for Linked Data is Resource Description Framework (RDF). It is the way to represent the data or resources on the Web. RDF is the most important concept to understand in order to understand Linked Data. In order to understand RDF, first we need to know URI (Universal Resource Identifier).

In a nutshell, URI defines a unique address for anything on the Internet, much like the post mail address for every home on the earth. That "anything" on the Internet not only includes physical entities such as apple or Da Vinci, but includes abstract concepts like love and peace also. Take for example: the URI, http://example.org/yongming-wang, is unique in the whole Internet and it refers to one of the authors of this article, Yongming Wang (Note that example.org is not a real domain name. It's a web convention to be used to demonstrate a website example. Anyone can use it for demonstration purpose). The above URI is also a URL. In other words, URL is one type of URI on the Web. Recently, the name of URI has been changed to IRI, short for International Resource Identifier.

So, what exactly is RDF? And what role does it play in Linked Data technology? RDF is a data model which is used to identify or describe things (also called entities) and their relationships on the Web. A RDF statement, also called a RDF triple, has three parts: Subject, Predicate, and Object. The subject and object designate two separate things, and the predicate describes the relationship between the subject and object. Its format goes like this:

<subject> <predicate> <object>

Here is an example: <Bob> <is a friend of> <Alice>. It can be expressed in a graph as seen in Figure 1.

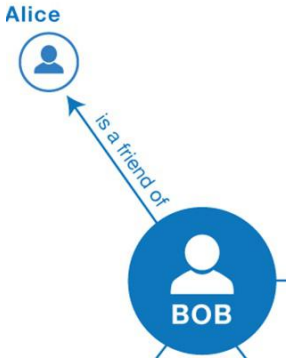

Figure 1: Part of Informal graph of the sample triples. Reprinted from W3C RDF 1.1 Primer[1]

[1]Retrieved from https://www.w3.org/TR/rdf11-primer/#fig1. Copyright © [24 June 2014] World Wide Web

According to the Linked Data principles, a subject must comprise Universal Resource Identifier (URI), and an object can be either a URI or a literal. The predicate should be defined by RDF Vocabulary and has to be a URI. RDF Vocabulary is like a schema in a relational database. The difference is that while a schema is to define the scope and type of the data, RDF Vocabulary is to define the relationship between data elements. Therefore, in the above example, the subject <Bob> is stated in the format like http://example.org/bob/ which is a unique URI on the Web, or a unique URL (HTTP URI). The object <Alice> is another unique URL such as http://example.org/alice. Lastly, the predicate <is a friend of> is also presented in the form of a URI such as http://example.org/friendOf/. To write the statement "<Bob> <is a friend> <Alice> explicitly in an official RDF statement, it should look like this:

<http://example.org/bob> <http://example.org/friendOf> <http://example.org/alice>

Multiple triples will be shown as a graph describing multiple things and their relationships. Furthermore, the beauty of Linked Data is that those multiple things and relationships reside across the Internet at different locations (aka web servers). The application developers can write applications to aggregate those things (data) on the fly. The users can follow those links (the relationships) to find more information relevant to their interest. See Figure 2 for an example.

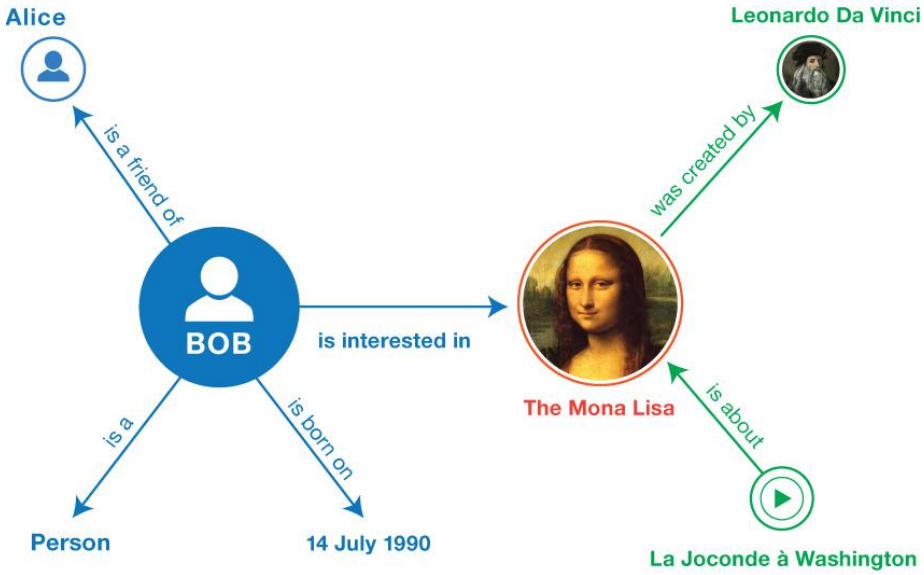

Figure 2: Informal graph of the sample triples. Reprinted from W3C RDF 1.1 Primer[2]

Consortium, (MIT, ERCIM, Keio, Beihang). http://www.w3.org/Consortium/Legal/2015/doc-license

[2] Retrieved from https://www.w3.org/TR/rdf11-primer/#fig1. Copyright © [24 June 2014] World Wide Web Consortium, (MIT, ERCIM, Keio, Beihang). http://www.w3.org/Consortium/Legal/2015/doc-license

There are six triples in Figure 2, in the form of <subject> <predicate> <object>. They are:
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>

This kind of graph can be extended unlimitedly. In such a way, almost everything on the Web can be uniquely described and linked together. Eventually, the graph will look like the giant graph in Figure 3.

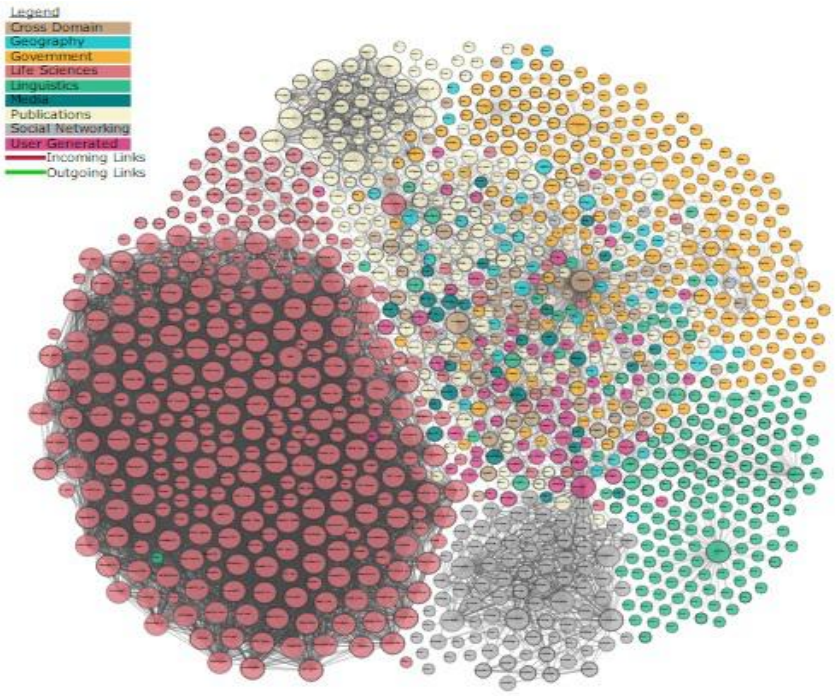

Figure 3: Linking Open Data cloud diagram 2017[3]

RDF Vocabularies

It will be a challenge to tackle the topic of RDF vocabulary as compared to other Linked Data concepts such as RDF data model that this paper talked about before and RDF serialization that will be discussed later. Much of the confusion and the slow adaptation of Linked Data is caused by the complexity of RDF vocabulary. Simply put, RDF vocabularies are like the schema in a relational database. According to Wood (2014):

[3] Reprinted from The Linking Open Data cloud diagram, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. Retrieved from http://lod-cloud.net/. CC-BY-SA license.

> "[RDF Vocabularies] provide definitions of the terms used to make relationships between data elements. Unlike a relational database's schema, however, RDF vocabularies are distributed over the Web and are developed by people all over the world, and only come into common use in Linked Data if a lot of people choose to use them." (p. 38)

RDF vocabulary itself is in HTTP URI format and is defined and preserved in the well-known web places. Examples follow. RDF vocabulary is used in the predicate position of the RDF triples to define the relationship between a subject and an object. Let us take the previous example of "Bob is a friend of Alice" to show how the RDF vocabulary is used.

As shown before, the RDF statement for "Bob is a friend of Alice" is written like this:

<http://example.org/bob> <http://example.org/friendOf> <http://example.org/alice>

To use the RDF vocabulary in the predicate position in real life, we need to replace <http://example.org/friendOf> with: http://xmlns.com/foaf/0.1/knows.

Let's take a look at it closely. There are basically two parts in http://xmlns.com/foaf/0.1/knows: http://xmlns.com/foaf/0.1/ is the namespace for FOAF (Friend of a Friend) Vocabulary, and "knows" part is a property of FOAF Vocabulary. FOAF is an open source project developed in mid-2000 for linking people on the Web. FOAF is widely used in social networking by many Linked Data projects. When we want to describe that someone is a friend of someone else, we can use this FOAF property and any computer program of Linked Data can automatically recognize and understand its meaning.

As mentioned above, the FOAF Vocabulary is in HTTP URI format (http://xmlns.com/foaf/0.1/), and it's kept at the well-known web place (http://xmlns.com).

A namespace is generally considered a placeholder to uniquely identify a set of names or properties. In the namespace of "foaf", short for http://xmlns.com/foaf/0.1/, all its properties, including the one we use here, "knows", are centrally preserved and uniquely defined. In other words, no ambiguity exists that "foaf:knows" (short for http://xmlns.com/foaf/0.1/knows) refers to a relationship between two persons. We will never confuse it with other "knows." In this sense, the namespace can also be called the prefix.

In order to fully understand RDF vocabularies, and especially, to be able to create your own RDF vocabulary, it is essential to learn the two key components of RDF vocabulary: Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL). RDFS is the definition language for RDF vocabulary. RDFS defines the classes and types which helps create new RDF vocabularies. OWL is an extension of RDF. Due to their complexities and the length limitation of this paper, the authors will not elaborate them here.

Another important RDF vocabulary is Simple Knowledge Organization System (SKOS). The main purpose of SKOS is to turn the traditional controlled vocabularies such as thesauri and all sorts of subject headings (e.g. Library of Congress Subject Heading) into RDF vocabularies. This feature makes SKOS especially important for the library community.

For better understanding of RDFS, OWL, and SKOS, the authors recommend a book titled "Semantic Web for the Working Ontologist" by Dean Allenmang and James Hendler (Allenmang & Hendler, 2012). As the book includes many examples and is written in an easy

and light style, Allenmang and Hendler makes learning the rather difficult topics of semantic modeling and ontology an easy task.

RDF Serialization

RDF Serialization, is the way the RDF statements are written so that the computer program can read and process them. There are different types of RDF serialization. The common ones are: Turtle (short for Terse RDF Triple Language), RDF/XML (the original RDF format in XML), RDFa (RDF embedded in HTML attributes), and the newer and more popular one called JSON-LD. This paper will focus on JSON-LD in this article.

JSON-LD, short for JavaScript Object Notation (JSON) for Linking Data, became popular because it's a favorite scripting language for many web developers and almost all the programming languages have multiple libraries to parse it. JSON is easy to write and read. Let's still take the previous example to show how its RDF triples can be written in JSON-LD format.

Let's take the following three RDF statements:

<Bob> <is a> <person>.

<Bob> <is a friend of> <Alice>.

<Bob> <is born on> <the 4th of July 1990>.

Their formal RDF triples are:

<http://example.org/bob> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>

<http://example.org/bob> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice>

<http://example.org/bob> <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date>

In the above example, http://www.w3.org/1999/02/22-rdf-syntax-ns#type can be shorten as rdf:type, which belongs to RDFS.

http://schema.org/birthDate comes from another popular RDF Vocabulary, schema.org. And 1990-07-04 is a literal object of the type Date as defined in XML Schema.

The JSON-LD format is as following:

```
{
  "@context": {
   "foaf": "http://xmlns.com/foaf/0.1/",
   "Person": "foaf:Person",
   "knows": {
    "@id": "foaf:knows",
    "@type": "@id"
   },
   "birthdate": {
    "@id": "http://schema.org/birthDate",
    "@type": "http://www.w3.org/2001/XMLSchema#date"
   }
  },
```

```
  "@id": "http://example.org/bob#me/",
  "@type": "Person",
  "birthdate": "1990-07-04",
  "knows": "http://example.org/alice#me/"
 }
```

As illustrated above, JSON-LD format is easy to understand. They are all in key-value pairs. The only tricky part is the context object inside which the prefixes or namespaces are defined.

SPARQL – The Query Language

SPARQL is not an acronym. Its whole name is SPARQL Protocol and RDF Query Language. SPARQL is the querying language for RDF dataset just as SQL is the query language for relational databases. The syntax of SPARQL and SQL are similar. But their similarity stops there. Actually, in order to learn SPARQL quickly, one should forget what one has learned about SQL.

We can use SPARQL to query the local RDF file with RDF data in the form of triples (see examples later). We can also use SPARQL to query remote RDF data store no matter where it is on the Web as long as that RDF data store provides a SPARQL endpoint service. Further, we can combine any number of local and remote queries to get the data we want in our application. That is the real power of SPARQL and Linked Data.

First, we will start with a simple SPARQL example. We will demonstrate how to use SPARQL to query a local RDF file. Suppose we have a file named bob.rdf with the following content:

```
(bob.rdf)
prefix foaf: http://xmlns.com/foaf/0.1/ .
prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns .
prefix schema: http://schema.org/ .
prefix xsd: http://www.w3.org/2001/XMLSchema# .

http://example.org/bob#me  rdf:type foaf:person .
http://example.org/bob#me  schema:birthdate “1990-07-04”^^xds:date .
http://example.org/bob#me  foaf:knows http://example.org/alice#me .
http://example.org/bob#me  foaf:knows http://example.org/lisa#me .
```

We want to use SPARQL to find all Bob’s friends in bob.rdf file. Here is the query:

(Notes: SPARQL finds the result by pattern matching. Any word with a question mark is a variable.)

```
prefix foaf: http://xmlns.com/foaf/0.1/ .

select ?x
from bob.rdf
```

```
where {
 http://example.org/bob#me foaf:knows ?y ;
}
```

The result will be like this:

x
alice
lisa

Inside the where clause, ?y is the variable. http://example.org/bob#me and foaf:knows are given values. It asks to find the value of the object position in all triples given the subject of http://example.org/bob#me and the predicate of foaf:knows. Hence alice and lisa. If we want to find both subject and object based on the predicate's value of foaf:knows, we will use the query as following:

```
prefix foaf: http://xmlns.com/foaf/0.1/ .

select ?x (as Name) ?y (as Friend)
from bob.rdf
where {
 ?a foaf:knows ?b ;
}
```

The result:

Name	Friend
bob	alice
bob	lisa

Once again, SPARQL finds the result by matching the given values in the where clause, and return all values for the variables in one or more triple positions.

This is just a simple introduction to SPARQL. It can get rather complicated in the real application.

INVOLVEMENT OF LIBRARY COMMUNITY

Libraries became aware of the value of Linked Data and the Semantic Web as a great way to describe library resources as early as 2005 when the US, Canada, and UK formed a joint committee to revise AACR2 cataloging rule. The release of RDA (Resource Description and Access) in 2010 provided guidelines to catalog and describe library resources in such a way that the resulting bibliographic data will be in alignment with Linked Data, a Web standard recognized and shared by other communities on the Internet. The advantages of Linked Data are

manifold, including release of bibliographic data from the silos to the web, link to resources from other communities, and retrieval of library resources by Internet search engines. According to research, 82% of the information consumers start by searching an Internet search engine and only 1% from a library's website (OCLC, 2011; Wordstream, 2017). Exposure of rich library data in the Semantic Web and the Internet will lead to more use of library resources and better services to users.

Since 2010 libraries are vigorously pursuing the goal of transforming bibliographic data into Linked Data which is the required format for the Semantic Web. The road is rocky and the development is slower than anticipated. One reason is that Linked Data is very new to libraries and it is a drastic departure from the traditional cataloging practice. Many technicalities need to be ironed out before the new practice is put to production. The lack of participation could be another possible setback. So far only big libraries and organizations have the technical expertise and financial resources to devote to the test and development of the Linked Data projects. LC, OCLC, and other national libraries have been the leading forces in Linked Data projects in libraries. Most small libraries are watching and waiting rather than participating. There is a lack of prototypes that will demonstrate the benefits of library data as linked data and many librarians still cannot envision how the future Semantic catalog looks and works. The magnitude of data involved, about 40 years of cataloged data, is not an easy task to be transformed into linked data. Library vocabularies and ontologies are complex and take a long time to complete. In spite of the aforementioned obstacles, Linked Data is the right path that libraries worldwide chose to follow and they have made great progress. The following is a description of the accomplishments by library community towards Linked Data and the Semantic Web.

Library of Congress (LC)

LC has been a world leader in promoting Linked Data technologies and their potential applications in libraries. The first move made by LC was to convert the LCSH, Name Authority File, and other controlled languages into RDF statements and URIs, and thus made them ready for use by other Semantic Web applications. LC is also instrumental in the development of RDA cataloging rule which is based on FRBR (Functional Requirement Requirements for Bibliographic Records) and supports Linked Data. After the release of RDA in 2011, LC immediately began its work on BIBFRAME (Bibliographic Framework), which is a new display standard intended to replace MARC. In late 2012 and early 2013 BIBFRAME 1.0 was released for testing in a pilot project. It included a series of tools such as BIBFRAME Editor, MARC to BIBFRAME Comparison Viewer, and MARCXML to BIBFRAME Transformation Tool. LC has been diligently testing and modifying BIBFRAME since then. This is a time consuming and complicated process. BIBFRAME ontologies or vocabularies are the core and also the more difficult part of BIBFRAME development. Conversion of existing MARC to BIBFRAME is another challenge. LC has 19 million MARC records (McCallum, 2017).

The latest data model and second generation of BIBFRAME is BIBFRAME 2.0 released in 2016. The revised new data model includes three core categories of abstraction: work, instance, and item and further defines three additional concepts related to the core categories: agents,

subjects, and events (LC, 2016). BIBFRAME 2.0 has released MARC to BIBFRAME Conversation Specifications and Programs. However, BIBFRAME Editor 2.0 is still under construction. The BIBFRAME Editor 2.0 will have more complete ontologies that have classes and properties specially designed to describe library resources. The two major vocabularies or ontologies used in BIBFRAME 2.0 are BIBFRAME and MADSRDF (Metadata Authority Description Schema in RDF). In addition, BIBFRAME 2.0 also draws on a few foundation ontologies developed by World Wide Web Consortium including OWL (Web Ontology Language), RDFS (RDF Schema), and SKOS (Simple Knowledge Organization System). "The BIBFRAME 2 ontology is much better integrated with the RDF environment, yet it is also more in synch with the RDA cataloguing rules even while staying rule agnostic" (McCallum, 2017, p.79).

Despite of the complex ontologies and vocabularies, BIBFRAME Editor itself is a simple tool that will turn the bibliographic data via a web-based input screen into RDF statements, one of the building blocks for Linked Data. The BIBFRAME 2.0 Conversion Programs are expected to be able to process a bigger number of MARC records and include fuller data from MARC records. It is unknown as to how and where BIBFRAME data will be searched and displayed.

LC has made great progress in BIBFRAME development. It is obvious that BIBFRAME will be an ongoing project with future revisions of vocabularies and version releases long after it is in production.

Online Computer Library Center (OCLC)

OCLC is another leading force in Linked Data research and projects in libraries. Most of the OCLC Linked Data projects revolves around Worldcat.org, a database of more than 400 million bibliographic records from more than 16,000 libraries (OCLC Linked Data Research, 2017). Collaborating with LC and other national libraries, OCLC has achieved remarkable success in this area.

The first publicly visible project OCLC undertook was to add Schema.org mark-up to its Worldcat.org records. Schema.org is created by major Internet search engines such as Google, Bing, and Yandex that provides combined requirements and specifications for any individual or organization to follow if they want to be searched and displayed as linked data. "With the addition of Schema.org mark-up to all book, journal and other bibliographic resources in WorldCat.org, the entire publicly available version of WorldCat is now available for use by intelligent Web crawlers, like Google and Bing, that can make use of this metadata in search indexes and other applications" (Murphy, 2012). As Schema.org vocabulary is more general in nature and not detailed enough to describe library resources, OCLC also led and participated in the effort to reconcile Schema.org vocabulary with BIBFRAME vocabulary and the development of bibliographic extension of Schema.org vocabulary (http://bib.schema.org/).

OCLC implemented Worldcat.org Works so all the manifestations of the same work are linked and displayed in a cluster using the OCLC FRBR work set algorithm. "The algorithm collects bibliographic records into groups based on author and title information from bibliographic and authority records" (OCLC Linked Data Research, 2017). The Internet search

engine standards are followed as "The WorldCat Work entity is based upon properties defined by the schema:CreativeWork type" (OCLC Developer Network, 2017). The advantage of gathering all formats of a work under its title is self-evident. As of July 2017, about 215 million work entities are available in Worlcat.org (OCLC Linked Data Research, 2017).

OCLC Persons is a similar project except it is about person entities. "WorldCat person entities connect related information about a person into a brief description that includes various formats of the person's name, creative works that the person has produced, and biographic sources of information about the person. As of July 2017, WorldCat persons include more than 117 million descriptions of authors, directors, musicians, and others, which have been mined directly from WorldCat. These entities were used in a Linked Data pilot program in which libraries used WorldCat persons in their regular workflows" (OCLC Linked Data Research, 2017).

Virtual International Authority File (VIAF) is another successful Linked Data project initiated by OCLC and several national libraries including LC, German National Library, and French National Library. Located at https://viaf.org/, VIAF is an international authority file based on authority data from a list of national libraries and maintained by OCLC. To summarize its function, "VIAF matches and links the authority files of national libraries. It then groups all authority records for a given entity into a merged "super" authority record that brings together the different descriptions for that entity" (OCLC Linked Data Research, 2017). VIAF API allows users to search authority data by keywords, name, title, and more and retrieve authority records and relationships between authority records. VIAF is under Open Data Commons Attributions License and any individual or organization can use it. "VIAF has been available as Linked Data since 2009 and is now one of the most widely used Linked Data resources published by the library community" (OCLC Linked Data Research, 2017).

OCLC and LC collaborated in developing FAST (Faceted Application of Subject Terminology), a general-purpose subject heading schema derived from Library of Congress Subject Headings (LCSH). The purpose of FAST is to create a simple to use and easy to understand faceted subject scheme than LCSH. The two subject headings are compatible and LCHS can be converted into FAST. Since 2011 FAST is available as Linked Data. FAST is known to be used by some national libraries and organizations for subject indexing and metadata description. According to OCLC, FAST is "one of the library domain's most widely used subject vocabularies" (OCLC Linked Data Research, 2017).

The successful Linked Data projects of WorldCat Works and WorldCat Persons entities, and Schema.org Markup "helped drive more than 74 million visits to WorldCat.org in 2016 and more than 17 million visits to local library catalogs around the globe" (COCL Linked Data Research, 2017).

Other US Library Linked Data Projects

The library Linked Data movement also comprises projects undertaken by Zepheira and many other academic libraries. Eric Miller, CEO of Zepheira, a Linked Data consulting company that developed BIBFRAME 1.0, advocated immediate action by libraries to publish their data on the

web so Internet search engines can search and display them on the top of the result page. Toward this end, Zepheira started the Libhub Initiative in 2014 and Library.Link Network in 2016. Partnering with vendors including EBSCO, SirsiDynix, and Innovative Interfaces, the Library.Link Network project involves a four-step process in which "Zepheira copies a partner library's catalog, converts records into the structured BIBFRAME format, and then hosts these BIBFRAME records in the Library.Link global, shared content distribution network designed for large-scale web ingest" and "Creative Commons licensing—requiring attribution to the library—is also added to each record, ensuring that service providers such as Google and Microsoft know where the data came from and what companies are allowed to do with it" (Enis, 2016). The final step is to publish library bibliographic data, events, hours, and staff information on the web. The initial participants include public libraries. The work is under progress.

Data conversion is a key component in Linked Data development for libraries. Colorado College is leading two projects. "One is to convert not only MARC but other data they hold in formats like MODS, Dublin Core, and other XML file formats to BIBFRAME RDF for access across these files. Another converts MARC records to BIBFRAME and then converts BIBFRAME to schema.org for sending to Google" (McCallum, 2017, p.83).

A few large academic libraries received grants from the Andrew W. Mellon Foundation for collaboration on Linked Data projects from 2014 to 2018 (LD4L Project Team, 2016). The partner universities include Cornell, Harvard, Columbia, Stanford, Princeton, and others. The projects supported by the grants include Linked Data for Libraries (LD4L), Linked Data for Libraries Labs (LD4L Labs), and Linked Data for Production (LD4P). The goals of those projects is to create an ontology compatible with BIBFRAME and other existing ontologies for describing local scholarly collections, to develop an open source semantic system to edit, search, and display scholarly resources, to test and pilot workflow in Linked Data technical services, and to create tools and guidelines for future work University of California Davis Library also piloted a BIBFLOW project to study the workflow. Those efforts extend LC's work on Linked Data and will benefit all libraries.

National Library of Medicine also actively participated in BIBFRAME test and ontology development. In 2014 NLM published beta versions of two of its datasets as Linked Data: PubChemRDF, containing information on the biological activities of small molecules and MeSH RDF, NLM's thesaurus of Medical Subject Headings. Both RDF products are searchable from their own SPARQL query interfaces or querying can be directly integrated into programs and services using their SPARQL endpoints (Davis Library, University of California, 2016).

Library System Vendors

BIBFRAME Editor 2.0 is still not released. Therefore, it is hard at this stage for library system vendor to invest money and manpower into a data model that is still evolving. However, some vendors expressed their commitment to Linked Data and their intention to incorporate BIBFRAME into their systems. A few have taken actions to prepare for the BIBFRAME Editor 2.0 release.

1. Ex Libris is developing what is called BIBFRAME publishing feature which will turn MARC into BIBFRAME data. The company's roadmap for Alma includes cataloging in BIBFRAME format and discovery of materials cataloged in all formats in Primo including those in Linked Data. Innovative Interfaces, Inc. and SirsiDynix partnered with Zepheira to add additional function into their existing system which will enable libraries to transform MARC into BIBFRAME data. They will also incorporate library location data to make the display location-sensitive for patrons. They will enhance their discovery layers to discover Linked Data and connect to outside resources for enriched content.

2. In September of 2017 librarians from 16 European countries and the US met in Germany and discussed the barriers for implementation of BIBFRAME. They felt that the lack of interest from the vendors of Integrated Library Systems (ILS) is one of the key issues. The discussion let to the publication of "BIBFRAME Expectations for ILS vendors" in February 2018 (Organizer Group 2018 European BIBFRAME Workshop, 2018).

Linked Data Projects in Non-US Libraries

Libraries in the world are watching closely the development of BIBFAME 2.0 and preparing themselves for the release of the new display standard. Libraries in Europe became interested in Linked Data and Semantic Web technologies long before the US libraries. European libraries are pioneers in Semantic Web technologies. The first known library catalog that embedded Linked Data is LIBRIS, the Swedish union catalog. As early as 2008 the catalog data became available as Linked Data and now it contains links to Wikipedia, DBpedia, LC authority files (names and subjects) and VIAF (Papadakis, Kyprianos & Stefanidakis, 2015).

The British Library began to publish its British National Bibliography (Linked Open BNB) as Linked Data as early as 2011. Statistics are not available as to how it is being used. French National Library (BNF) has been engaged in the project called "data.bnf.fr" which aims to make the catalog data of BNF into Linked Data. The goal of the project is to allow users to access library data on the web and link BNF data to DBpedia, VIAF, and other sources (Papadakis, Kyprianos & Stefanidakis, 2015. German National Library (DNB) is developing a Linked Data service for the long term commitment to Semantic Web and has been supplying its data in the RDF standard since 2010. The National Library of Spain (BNE) had a similar project called "datos.bne.es" which aims to release its bibliographic data as Linked Data and eventually to become part of the Semantic Web.

Canadian Linked Data Initiative (CLDI) is a collaboration between five Canada's largest research libraries, including National Library of Canada (Library and Archives Canada) and Bibliothèque et Archives Nationales du Québec. The participating libraries felt they were behind in many areas for the impending shift from MARC to Linked Data and BIBFRAME. The aim of the initiative is to get Canadian libraries up to date in strategic planning to embrace the changes in bibliographic control. The participants are discussing staff training, data preparation, enhanced discovery process and anything that is necessary to get Canadian libraries for a smooth transition into Linked Data world.

The Japanese National Library, also called National Diet Library (NDL), provides

metadata as Linked Open Data (LOD) to facilitate effective use by computer systems or applications. National Library of China (NLC) is vigorously engaged in research and discussions on Linked Data and Semantic Web technologies in Chinese language environment.

CONCLUSION

It has been almost 20 years since the inception of FRBR, then RDA, and now long waited BIBFRAME. The road to Linked Data has been bumpy, but there is no way to turn back. BIBFRAME will be an on-going development even with the upcoming release of BIBFRAME Editor 2.0. We hope that in the next five to ten years, most library data, including millions of bibliographic records in silos, will appear as Linked Data, freely and openly searchable and accessible on the web as many national libraries have done so. Yet libraries still face the new challenge to get bibliographic data into the search path of Internet search engines. "With an imperative to support novel means of discovery, and a wealth of experience in producing high-quality structured data, libraries are natural complementors to Linked Data" (Heath, 2011, p.36). What libraries are trying to accomplish will benefit the society. With that goal in mind, we will succeed. "The library community is poised to make great strides with semantic web technologies, as evidenced by recent endeavors involving BIBFRAME, a protocol that is largely considered to be the next generation standard for assigning and managing bibliographic metadata" (Johnson, 2015, p.42).

References

Allemang, D. & Hendler, J. (2012). *Semantic Web for the Working Ontologist. 2nd Edition.* Waltham, MA: Elsevier.

Banerjee, K. (2017). Translating Technobabble: All You Really Need to Know about URIs, Linked Data, and FRBR. *Computers in Libraries*, *37*(10), 21-24

Berners-Lee, T., & Fischetti, M. (2000). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web.* San Francisco: HarperBusiness.

Berners-Lee, T. (2006). Linked Data. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html,

Coyle, K. (2012). Linked Data Tools: Connecting on the Web. *Library Technology Reports, 48*(4). http://dx.doi.org/10.5860/ltr.48n4

Davis Library, University of California. (2016). *Survey of Current Library Linked Data Implementation*. Retrieved from https://bibflow.library.ucdavis.edu/xi-survey-of-current-library-linked-data-implementation/

Enis, M. (2016, June 21). Library.Link builds open web visibility for library catalogs, events. Retrieved March 27, 2018, from Library Journal website: http://lj.libraryjournal.com/2016/06/marketing/

library-link-builds-open-web-visibility-for-library-catalogs-event

Glaser, H., Millard, I., Sung, W., Lee, S., Kim, P., & You, B. (2009). Research on Linked Data and Co-reference Resolution. *International Conference on Dublin Core and Metadata Applications,* 0, pp. 113-117. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/958/957

Godly, C. J., & Smith-Yoshimura, K. (2017). From Records to Things: Managing the Transition from Legacy Library Metadata to Linked Data. *Bulletin of the Association for Information Science and Technology*, *43*(2), 18-23

Heath, T. & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space.* Morgan & Claypool Publishers.

Jin, Q., Hahn, J., & Croll, G. (2016). BibFrame Transformation for Enhanced Discovery. *LRTS*, *60*(4), 223-235.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC Horizon Report: 2015 Library Edition*. Austin, Texas:The New Media Consortium. Retrieved from https://www.nmc.org/publication/nmc-horizon-report-2015-library-edition/

LD4L Project Team. (2016). LD4L gateway. Retrieved March 27, 2018, from LD4L-Linked Data for Libraries website: https://www.ld4l.org/

Library Linked Data Incubator Group. (2011, October 25). *Library Linked Data Incubator Group Final Report*. Retrieved from https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

Library of Congress. (2016, April 21). *Overview of the BIBFRAME 2.0 Model*. Retrieved from https://www.loc.gov/bibframe/docs/bibframe2-model.html

McCallum, S. (2017). BIBFRAME Development [PDF]. *JLIS.it: Italian Journal of Library, Archives, and Information Science*, *8*(3). https://doi.org/10.4403/jlis.it-12415

Miles, A., Matthews, B., Wilson, M., & Brickley, D. (2005). SKOS Core: Simple knowledge organisation for the Web. *International Conference on Dublin Core and Metadata Applications,* 0, pp. 3-10. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/798

Mitchell, E. T. (2013). Library Linked Data: Research and Adoption. *Library Technology Reports. 49*(5). http://dx.doi.org/10.5860/ltr.49n5

Mitchell, E. T. (2016). Library Linked Data: Early Activity and Development. Library Technology Reports. 52(1). http://dx.doi.org/10.5860/ltr.52n1

Murphy, B. (2012, June 20). *OCLC Adds Linked Data to WorldCat.org*. Retrieved from https://www.oclc.org/en/news/releases/2012/201238.html

OCLC. (2013). *Meeting the E-Resources Challenge: An OCLC report on effective management, access and delivery of electronic collections* [PDF]. Retrieved from https://www.oclc.org/content/dam/oclc/reports/pdfs/OCLC-E-Resources-Report-US.pdf

OCLC Linked Data Research. (2017). *OCLC and Linked Data* [PDF]. Retrieved from https://www.oclc.org/content/dam/oclc/services/brochures/215912_WWAE-OCLC-Linked-Data-Report.pdf

OCLC Developer Network. (2017). *WorldCat Work Descriptions*. Retrieved from

https://www.oclc.org/developer/develop/linked-data/worldcat-entities/worldcat-work-entity.en.html

Organizer Group 2018 European BIBFRAME Workshop. (2018, February 8). *BIBFRAME Expectations for ILS Vendors*. Retrieved from https://wiki.dnb.de/display/EBW/Documents+and+Results

Papadakis, L., Kyprianos, K., & Stefanidakis, M. (2015). Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine*, *21*(5/6). https://doi.org/10.1045/may2015-papadakis

Szeto, K. (2017). The Mystery of the Schubert Song: The Linked Data Promise. *Notes. 74*(1). 9-23. http://dx.doi.org/10.1353/not.2017.0071

Tharani, K. (2015). Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BibFrame. *Information and Library Technologies*. *34*(1). https://doi.org/10.6017/ital.v34i1.5664

Wood, D., Zaidman, M., Ruth, L., & Hausenblas, M. (2014). *Linked Data: Structured Data on the Web.* Shelter Island, NY: Manning Publications Co.

Wordstream. (n.d.). Google Ads: What Are Google Ads & How Do They Work? Retrieved from http://www.wordstream.com/google-ads

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017). Representing and Integrating Bibliographic Information into the Semantic Web: A Comparison of Four Conceptual Models. *Journal of Information Science, 43*(4), 525-553.

About the authors

Yongming Wang is Systems Librarian / Associate Professor of The College of New Jersey. His research interests include Linked Data, Next-Gen library system, text and data analytics, digital library and institutional repository.

Sharon Q. Yang is Systems Librarian / Professor of Rider University. Her research interests include Linked Data and Semantic Web, library system and discovery service, institutional repository, open access, copyright.